



Computational Modeling and Analysis to Predict Intracellular Parasite Epitope Characteristics Using Random Forest Technique

Amir JAVADI^{1,2}, Ali KHAMESIPOUR³, Farshid MONAJEMI⁴, *Marjan GHAZISAEEDI¹

1. Department of Health Information Management, School of Allied Medical Sciences, Tehran University of Medical Sciences, Tehran, Iran
2. Department of Medical Social Sciences, Faculty of Medicine, Qazvin University of Medical Sciences, Qazvin, Iran
3. Center for Research and Training in Skin Diseases and Leprosy, Tehran University of Medical Sciences, Tehran, Iran
4. Ministry of Health and Medical Education, Tehran, Iran

***Corresponding Author:** Email: ghazimar@tums.ac.ir

(Received 10 Aug 2018; accepted 12 Oct 2018)

Abstract

Background: In a new approach, computational methods are used to design and evaluate the vaccine. The aim of the current study was to develop a computational tool to predict epitope candidate vaccines to be tested in experimental models.

Methods: This study was conducted in the School of Allied Medical Sciences, and Center for Research and Training in Skin Diseases and Leprosy, Tehran University of Medical Sciences, Tehran, Iran in 2018. The random forest which is a classifier method was used to design computer-based tool to predict immunogenic peptides. Data was used to check the collected information from the IEDB, UniProt, and AAindex database. Overall, 1,264 collected data were used and divided into three parts; 70% of the data was used to train, 15% to validate and 15% to test the model. Five-fold cross-validation was used to find optimal hyper parameters of the model. Common performance metrics were used to evaluate the developed model.

Results: Twenty seven features were identified as more important using RF predictor model and were used to predict the class of peptides. The RF model improves the performance of predictor model in comparison with the other predictor models ($AUC \pm SE: 0.925 \pm 0.029$). Using the developed RF model helps to identify the most likely epitopes for further experimental studies.

Conclusion: The current developed random forest model is able to more accurately predict the immunogenic peptides of intracellular parasites.

Keywords: Computational model; Immunogenic peptides; Intracellular parasites

Introduction

Historically, the most effective public-health prevention against infectious disease is vaccination. Development of an effective vaccine against any disease is a major breakthrough to control the disease. There have been tremendous efforts to de-

velop vaccines against infectious and non-infectious diseases, but yet no vaccine is available against many infectious diseases. The development of a new vaccine from theory to practice is a complex process task. Preclinical studies to de-